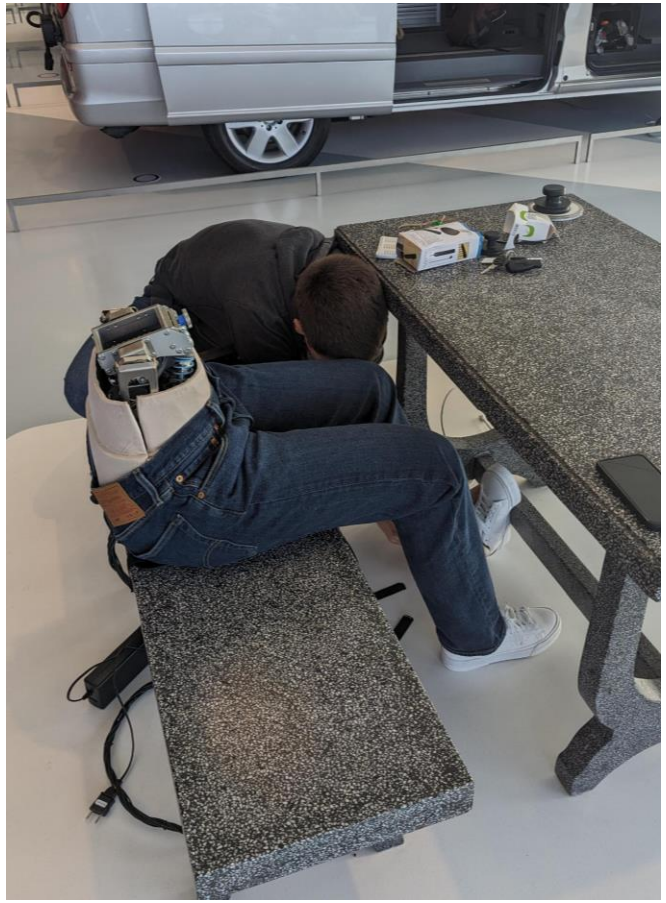
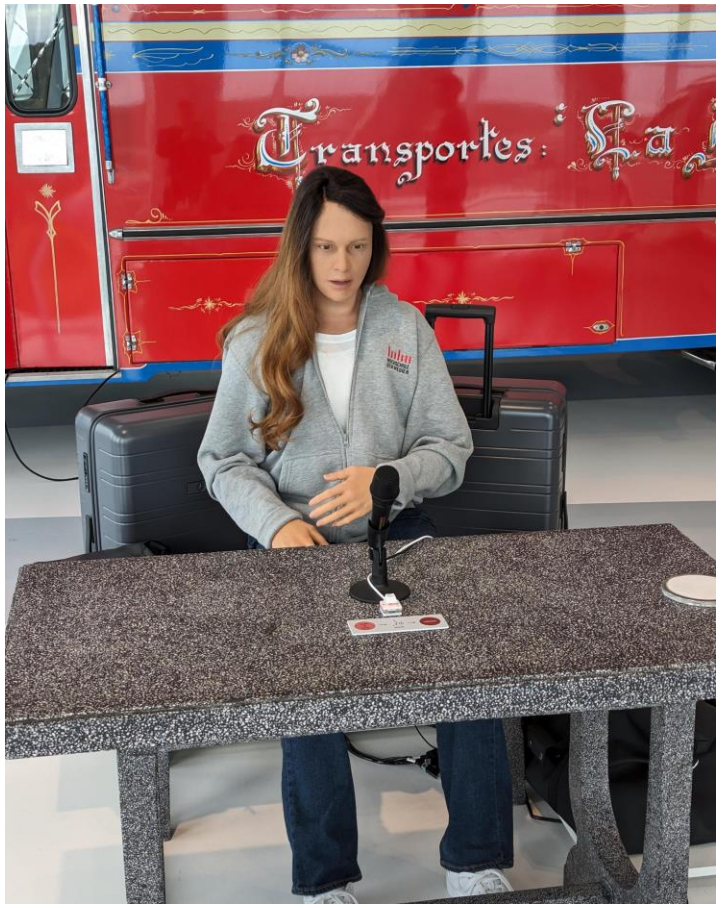


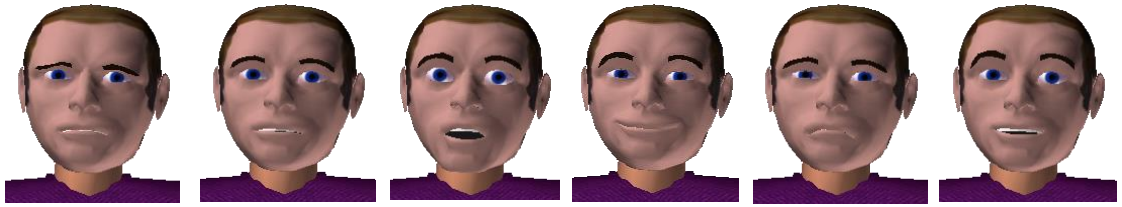


HOCHSCHULE DER MEDIEN

Emotion Simulation and Android Andrea in the wild

Waseda University, 22.05.2025

Prof. Dr. C. Becker-Asano, Hochschule der Medien, Stuttgart



WASABI emotion simulation

WASABI Affect Simulation for Agents with Believable Interactivity

WASABI – historical & theoretical roots

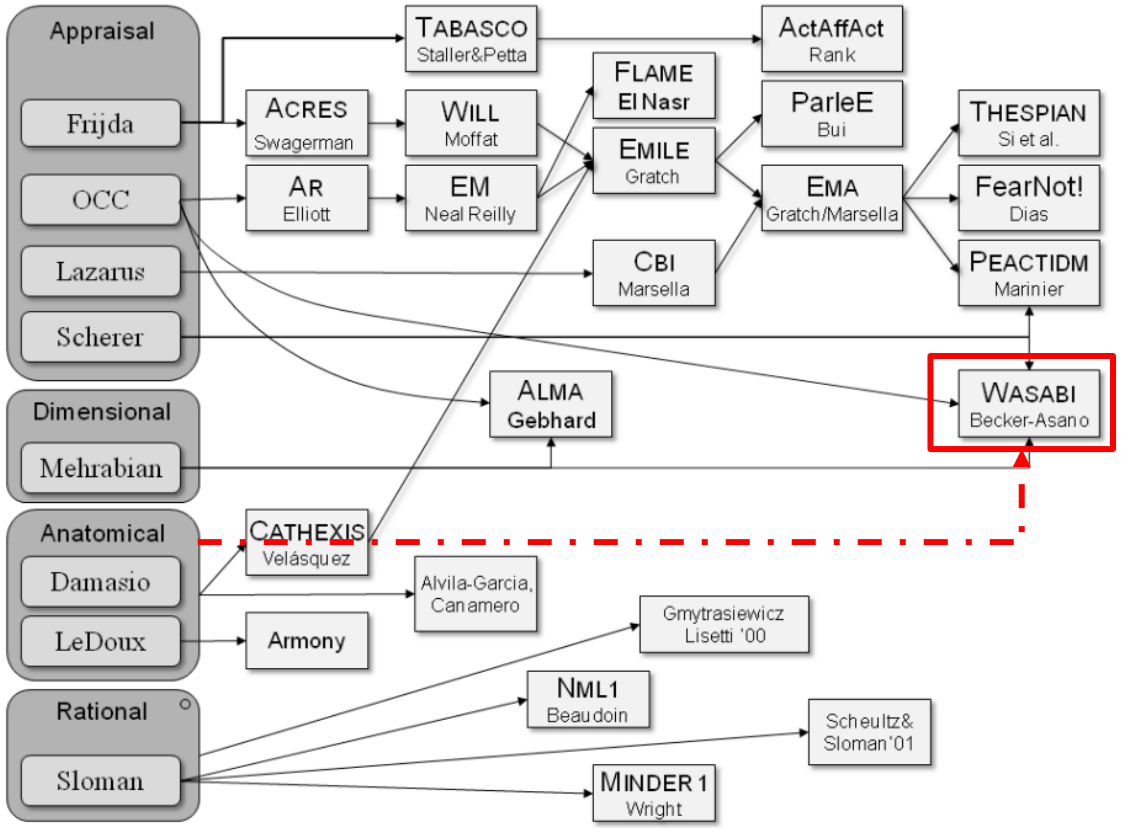
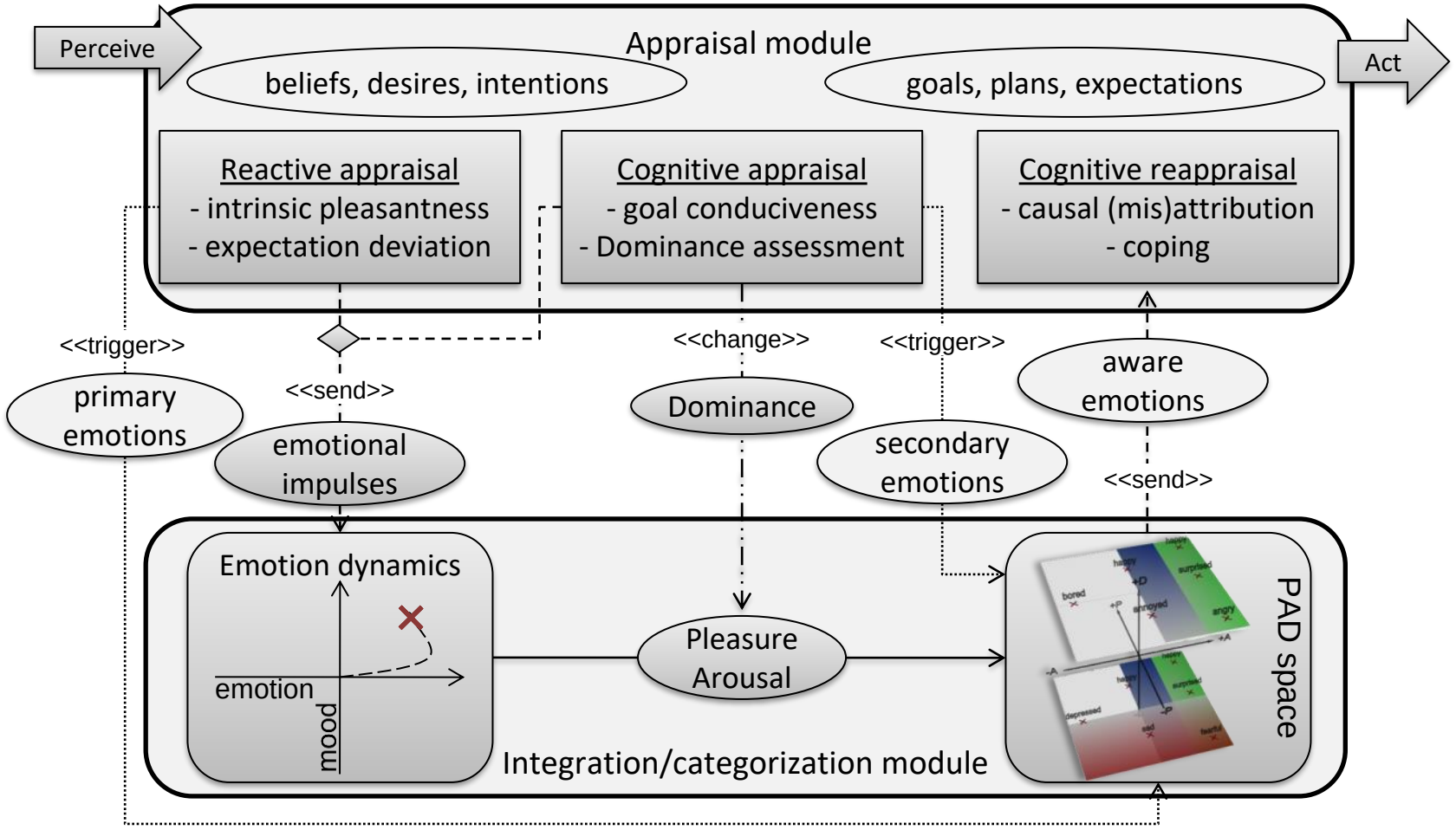


Figure 1: A history of computational models of emotion

- Development began in 2002 with the Multimodal Assembly eXpert MAX
- Inspired by seminal work of German philosopher and psychologist Wilhelm Wundt
- Dynamics of emotional experience is at its core
- Combines appraisal (OCC) with dimensional (PAD) emotion theories
- Applied to virtual humans and social robots
- „WASABI“ because „TABASCO“ was already taken ☺

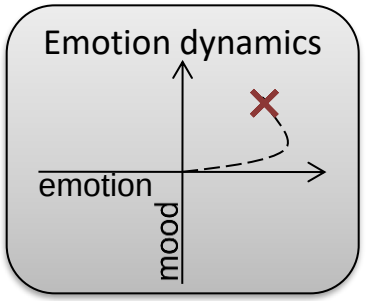
Source: Marsella, S., Gratch, J., & Petta, P. (2010). Computational models of emotion. A Blueprint for Affective Computing-A sourcebook and manual, 11(1), 21-46.

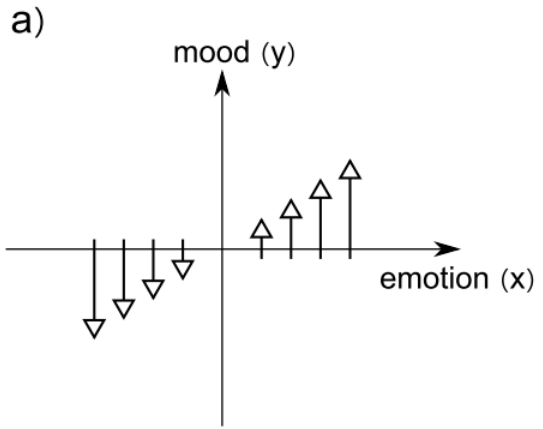
A short introduction to WASABI

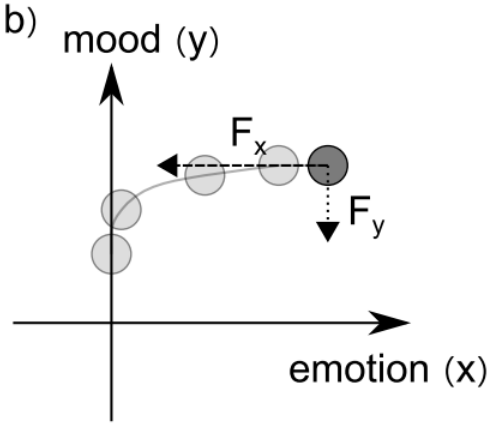


[W]ASABI [A]ffect [S]imulation for [A]gents with [B]elievable [I]nteractivity

WASABI





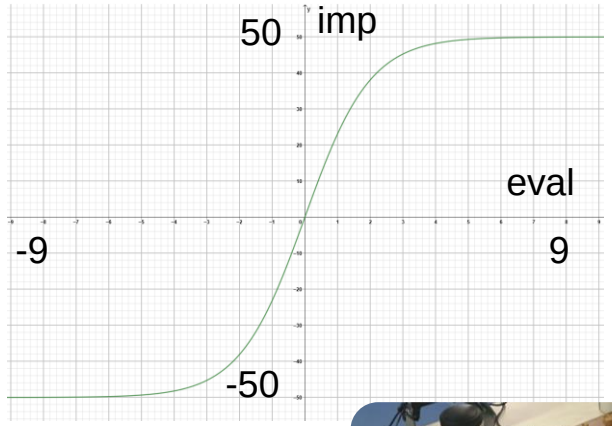


WASABI emotion dynamics:

- emotion combined with mood
- "valenced impulses" drive agent's emotional state,
  for example, with a chess engine:
  - with e_t the board evaluation:

$$imp(\mathrm{eval}_t) = k \times \tanh \frac{\mathrm{eval}_t}{r}$$

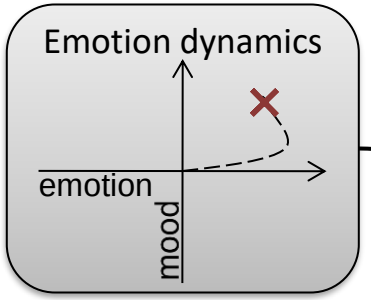
  - r reduces skewness of hyperbolic tangent (2)
  - k is a scaling factor (50)
  - E.g., $\mathrm{imp}(\mathrm{eval}_t) = 50 \times \tanh \frac{\mathrm{eval}_t}{2}$



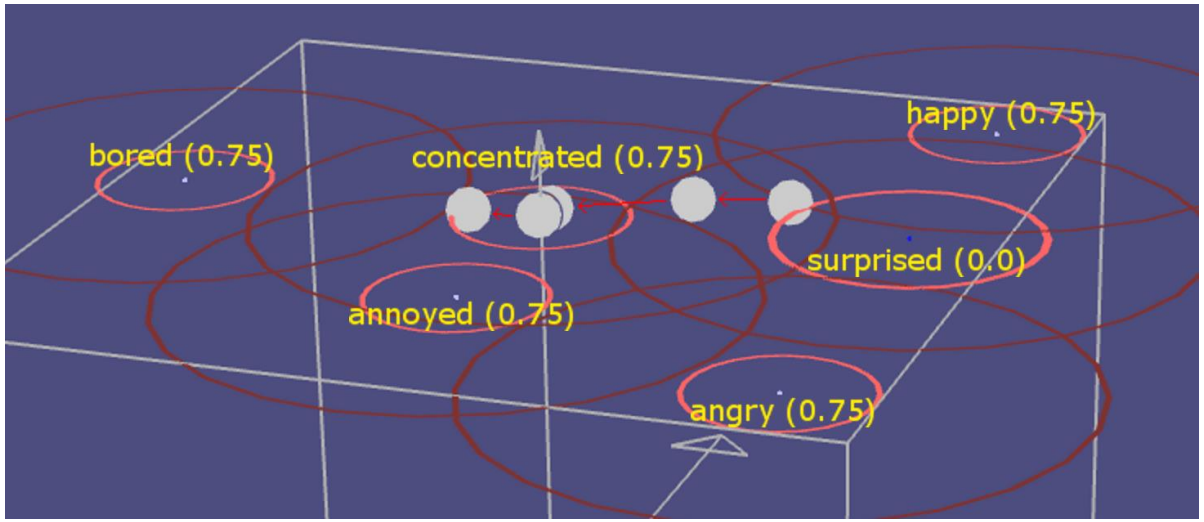
WASABI

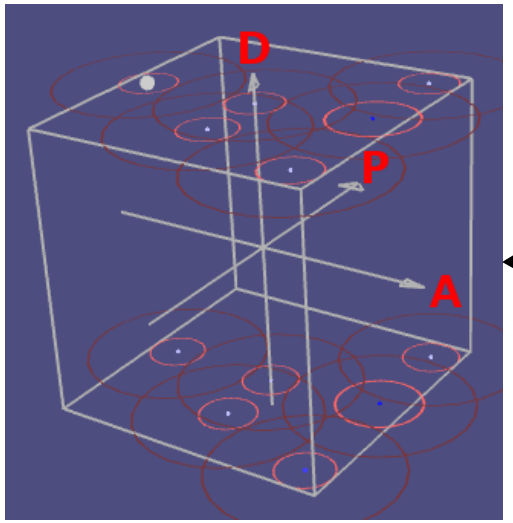
$$p(x_t, y_t) = \frac{1}{2} \cdot (x_t + y_t)$$

$$a(x_t, z_t) = |x_t| + z_t$$



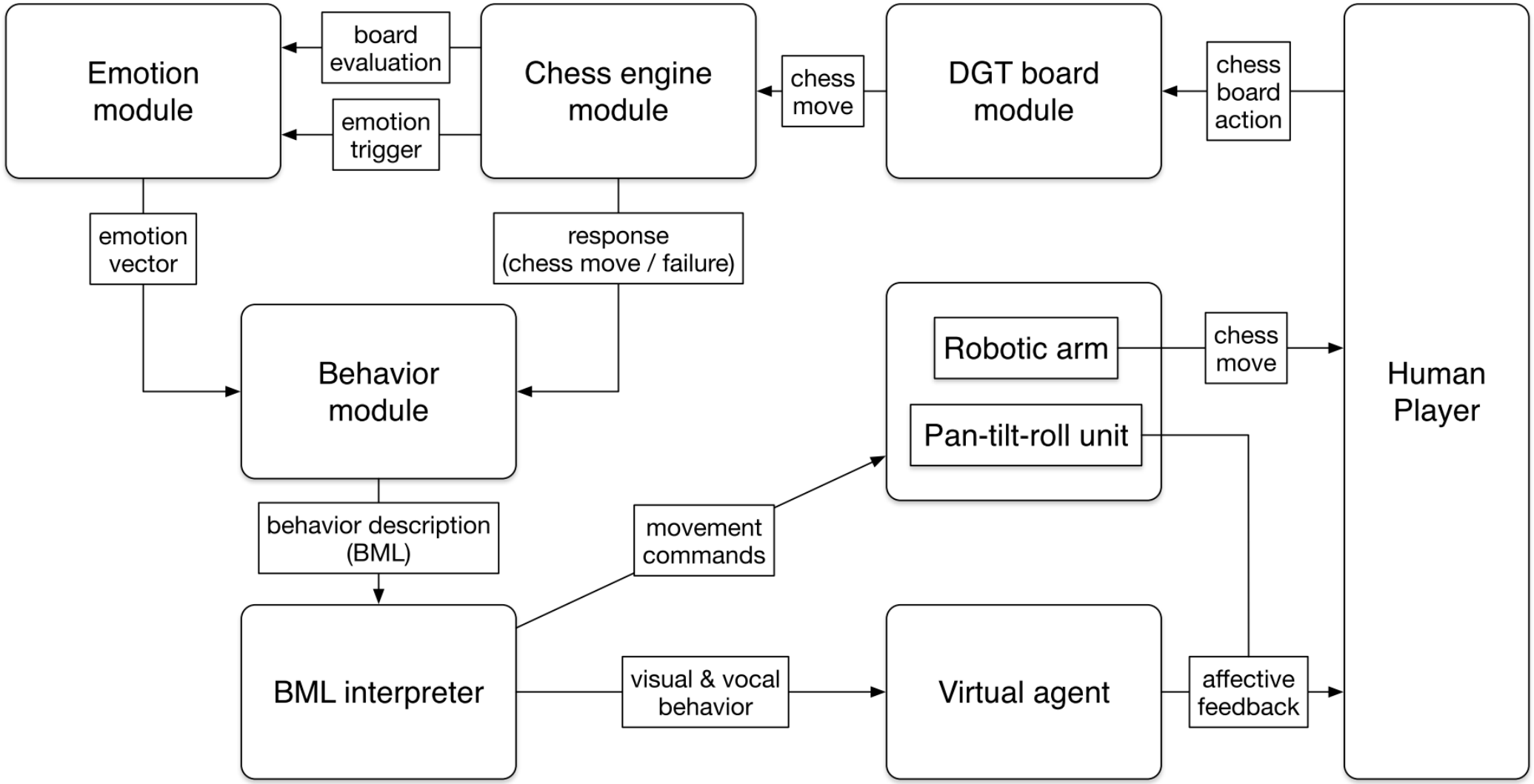
- Emotion dynamics in pleasure (P), arousal (A), and dominance (D) space
- Simple, UDP-based network interface
- Open source (hosted on GitHub) (→ https://github.com/CBA2011)
- (Presumably) quite well documented ☺





Software architecture of MARCO

2014

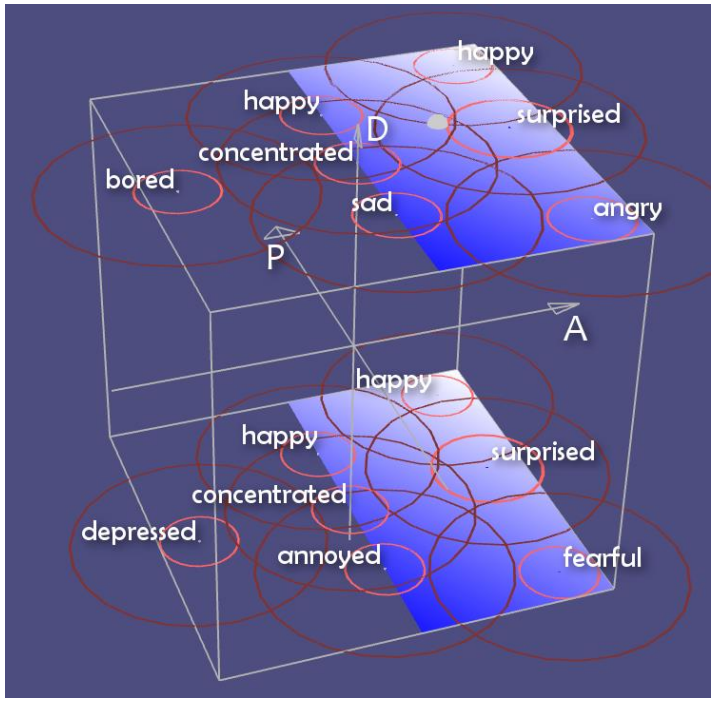


WASABI integration (2)

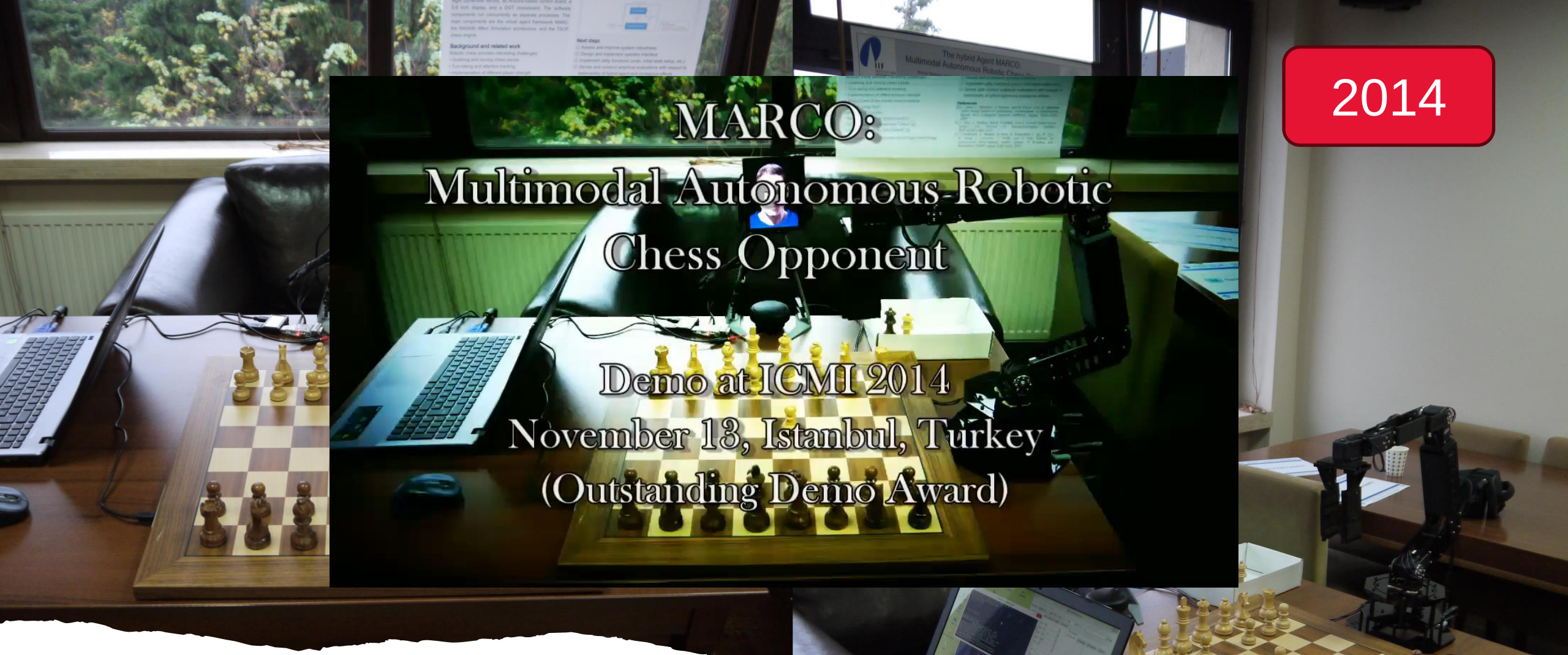
2014

HOCHSCHULE DER MEDIEN

- prospect-based emotions need to be triggered based on changing evaluations over time
- with e_t representing future directed evaluation:
  - fear: significant drop of board evaluation
  - surprise: significant change of board evaluation
  - fear-confirmed: if fear is present and situation as bad as expected (or even worse)
  - relief: if fear and situation much better than expected
  - hope: if good move at depth d of the search tree



trigger	*if..*
fear	$e_{t-1} - e_t > \epsilon$
surprise	$\lvert e_{t-1} - e_t \rvert > \epsilon$
fears-confirmed	$fear_{t-1} \wedge (e_{t-1} - e_t < \epsilon)$
hope	$e_{t,d} - e_{t,d-2} > \epsilon$
relief	$fear_{t-1} \wedge (e_t - e_{t-2} < \epsilon)$



2014

Outstanding demo award
ICMI 2014, Istanbul

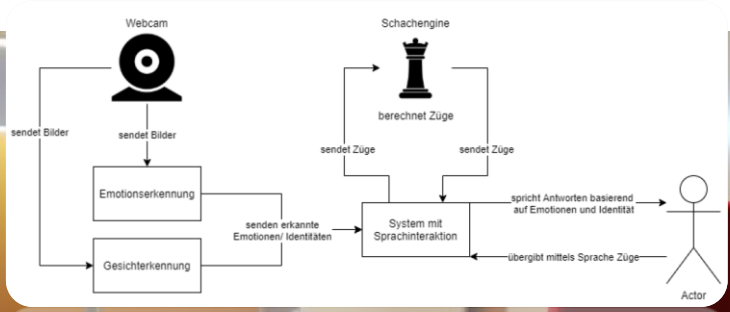
https://becker-asano.de/index.php/videos/robotics/87-marco-videos

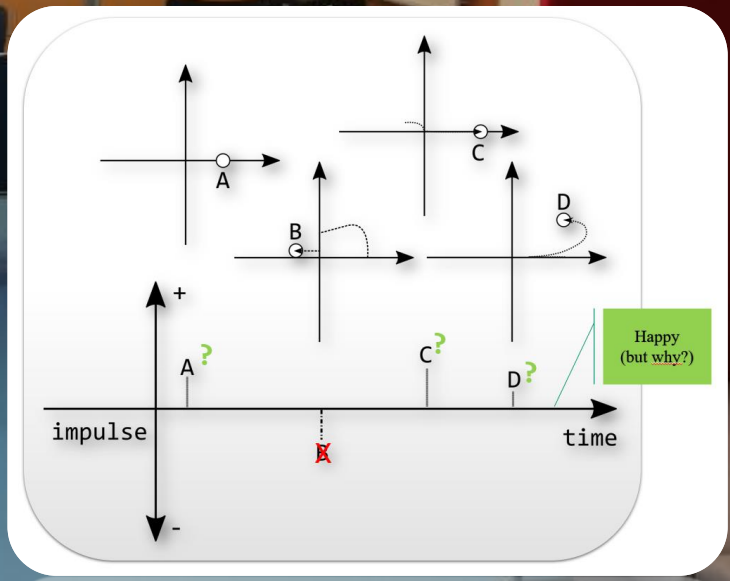
MARCO Bachelor & master thesis @ humanoidlab

- 2021, Sinan Kale, MA: „Development and integration of an **optical chess board recognition from an oblique angle** for the hybrid agent MARCO“
- 2021, Jannik Schmidt, BA: „Personalized spoken language interaction with a chess robot by **camera-based face and emotion recognition**“
- 2022, Florian Rapp, BA: „**Emotion detection from the speech signal**“
- 2022, Patrick Thomasius, BA: „**Misattribution of emotions** – analysis and simulation following an emotion dynamics approach“
- …

2020 →

HOCHSCHULE DER MEDIEN





Now to something completely different:

Android Andrea

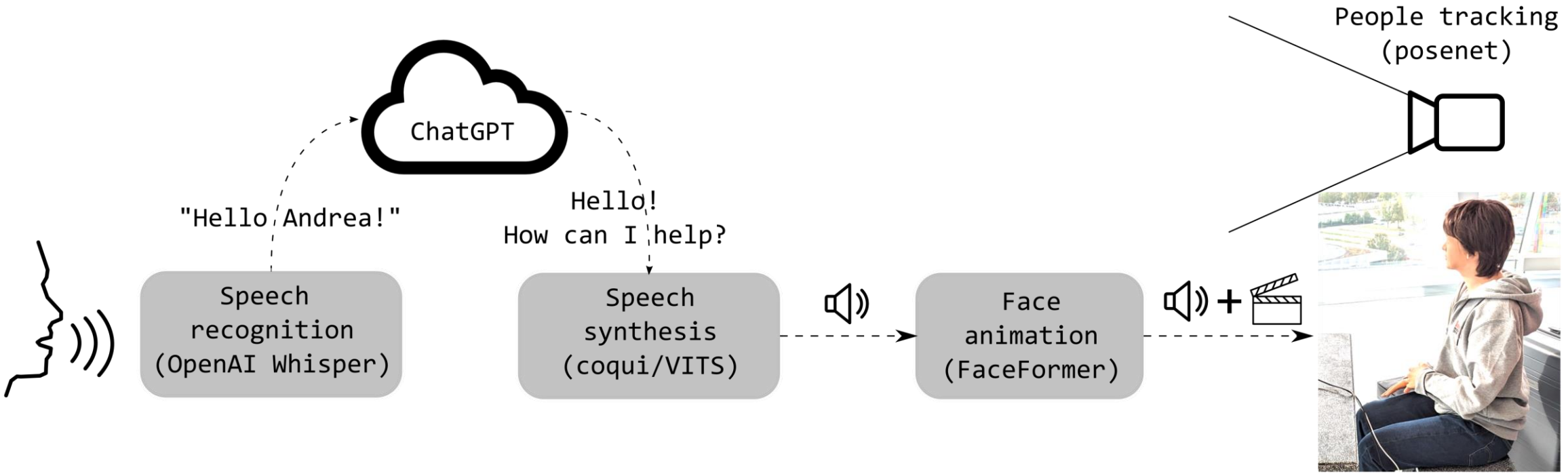
Impression of the setup

Project overview

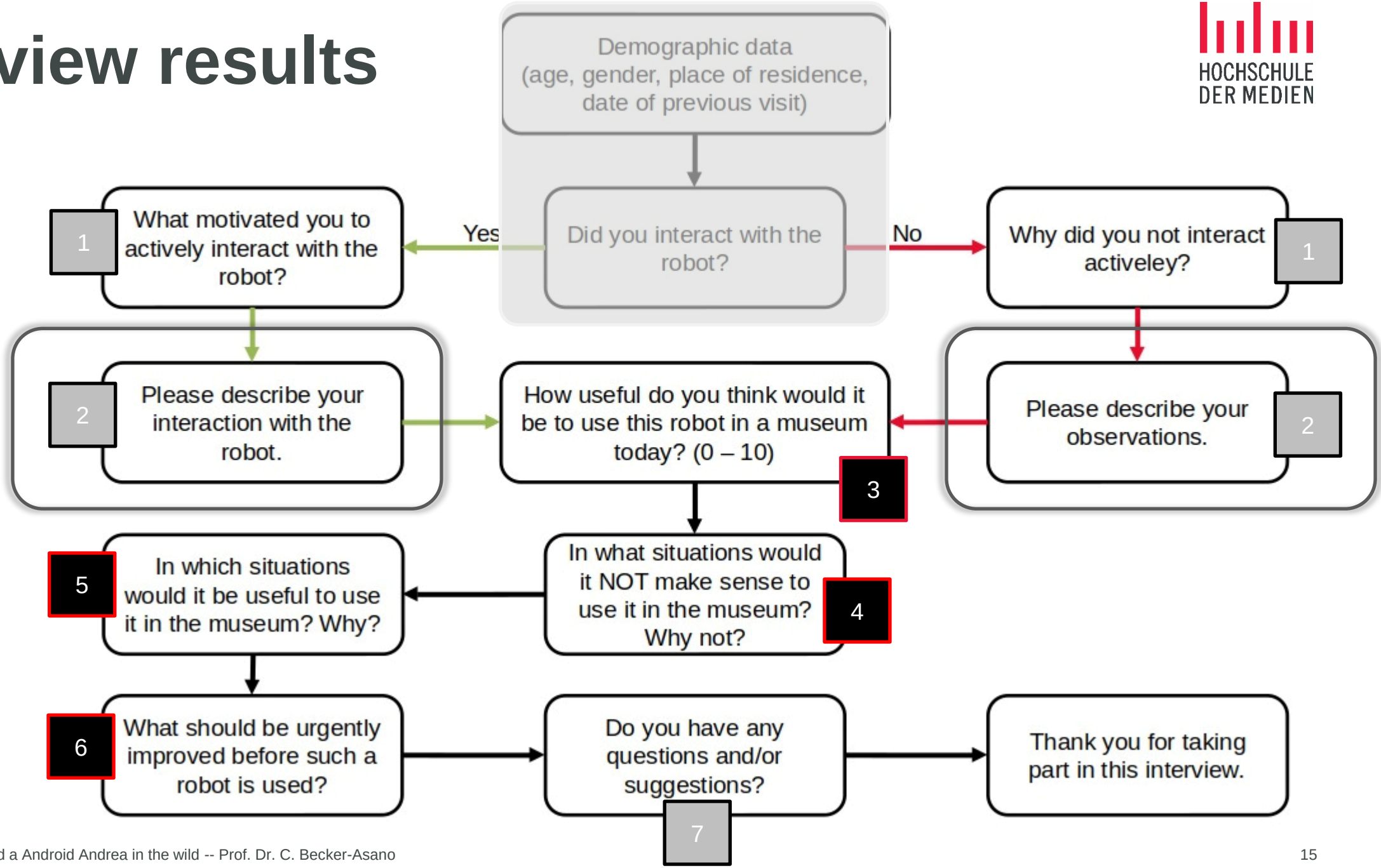
- Interactive installation of the android robot “Andrea” in the “Galerie der Reisen” of the Mercedes-Benz museum
- Daily from 9 am to 6 pm, October 31st until November 5th 2023
- Structured interviews with 44 randomly selected visitors
- Three versions of Andrea:
  1. Long hair with female voice (long-female, 2 days)
  2. Short hair with male voice (short-male, 3 days)
  3. Short hair with female voice (short-female, 1 day)
- Analysis of 44 interviews and 4436 audio input requests (transcribed to text) with corresponding system responses

Technical setup

HOCHSCHULE DER MEDIEN



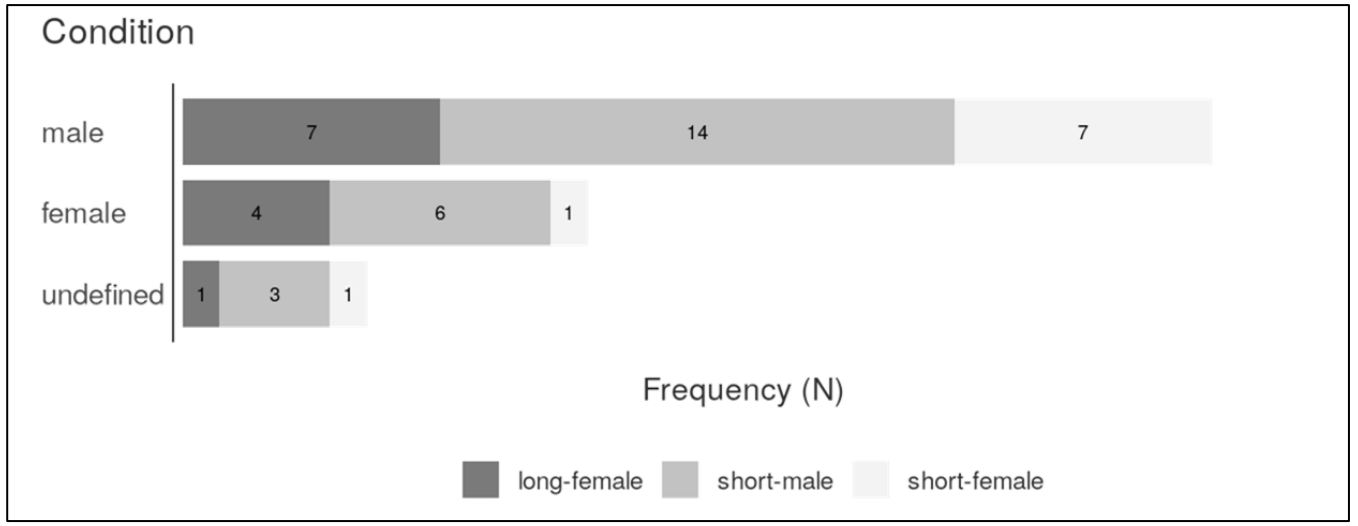
Interview results

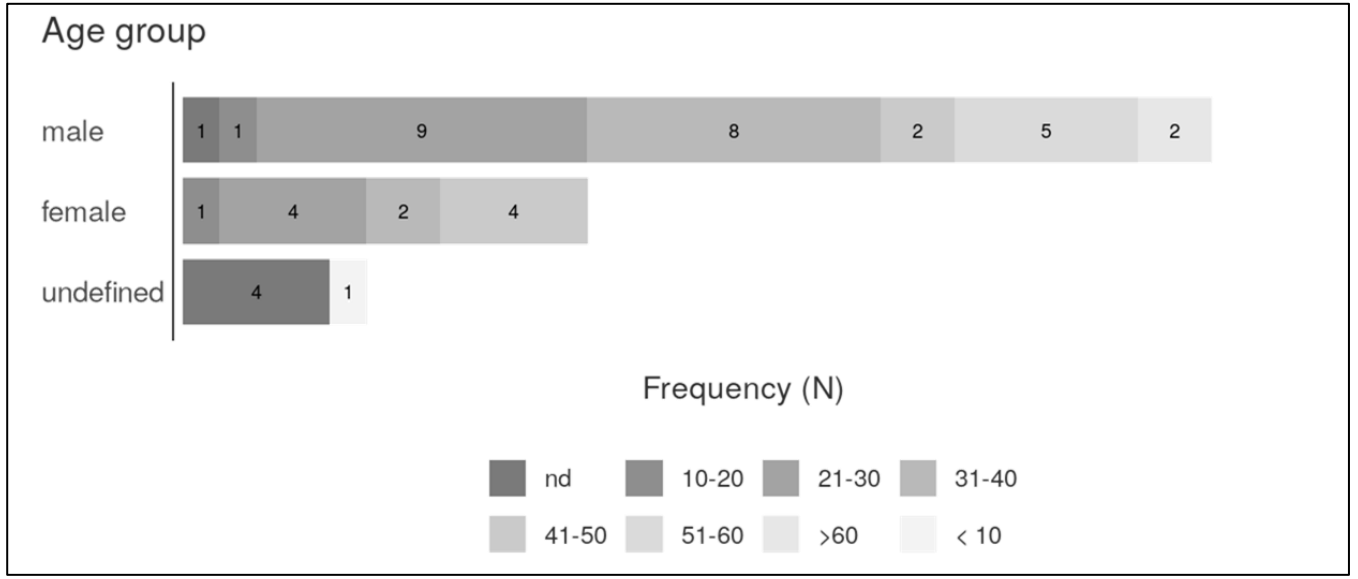


Gender over conditions / age groups

Demographic data
(age, gender, place of residence, date of previous visit)

- 28 male, 11 female, 5 undefined (missing data)
- Equal distribution of gender over experimental conditions
- Equal distribution of gender over age groups

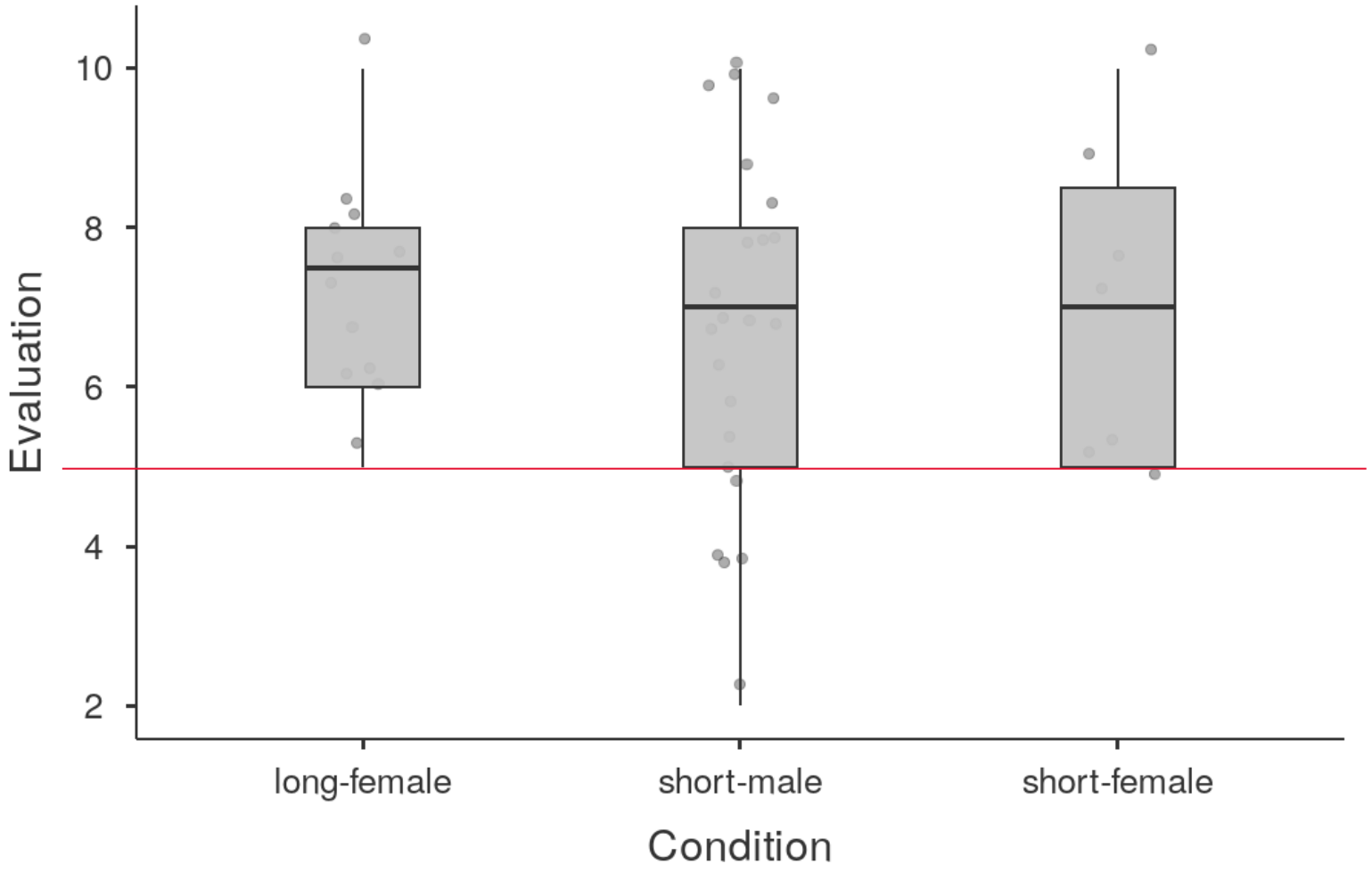




Usefulness on a 0 to 10 scale

HOCHSCHULE DER MEDIEN

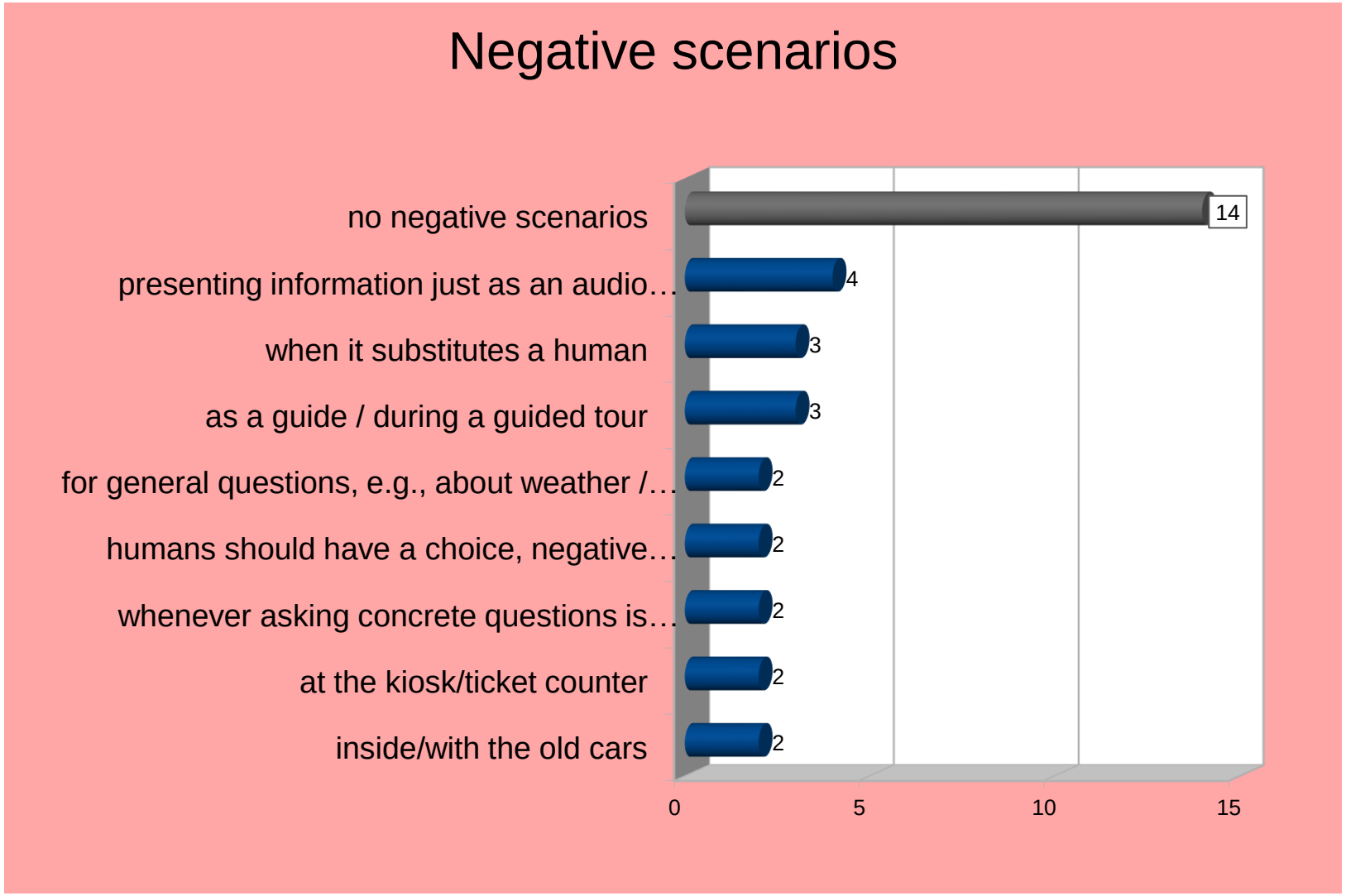
- All conditions well above average evaluation of five
  - pooled evaluation data: Shapiro-Wilk normality test $p = 0.056 \rightarrow$ Wilcoxon test with $H_0\ \mu \neq 5, p < 0.001$
- No significant differences between conditions
  - One-way ANOVA assuming unequal vairances with „Condition“ as grouping variable: $F = 0.237, df_1 = 2, df_2 = 15.9, p > 0.79$



Scenarios that are not making sense

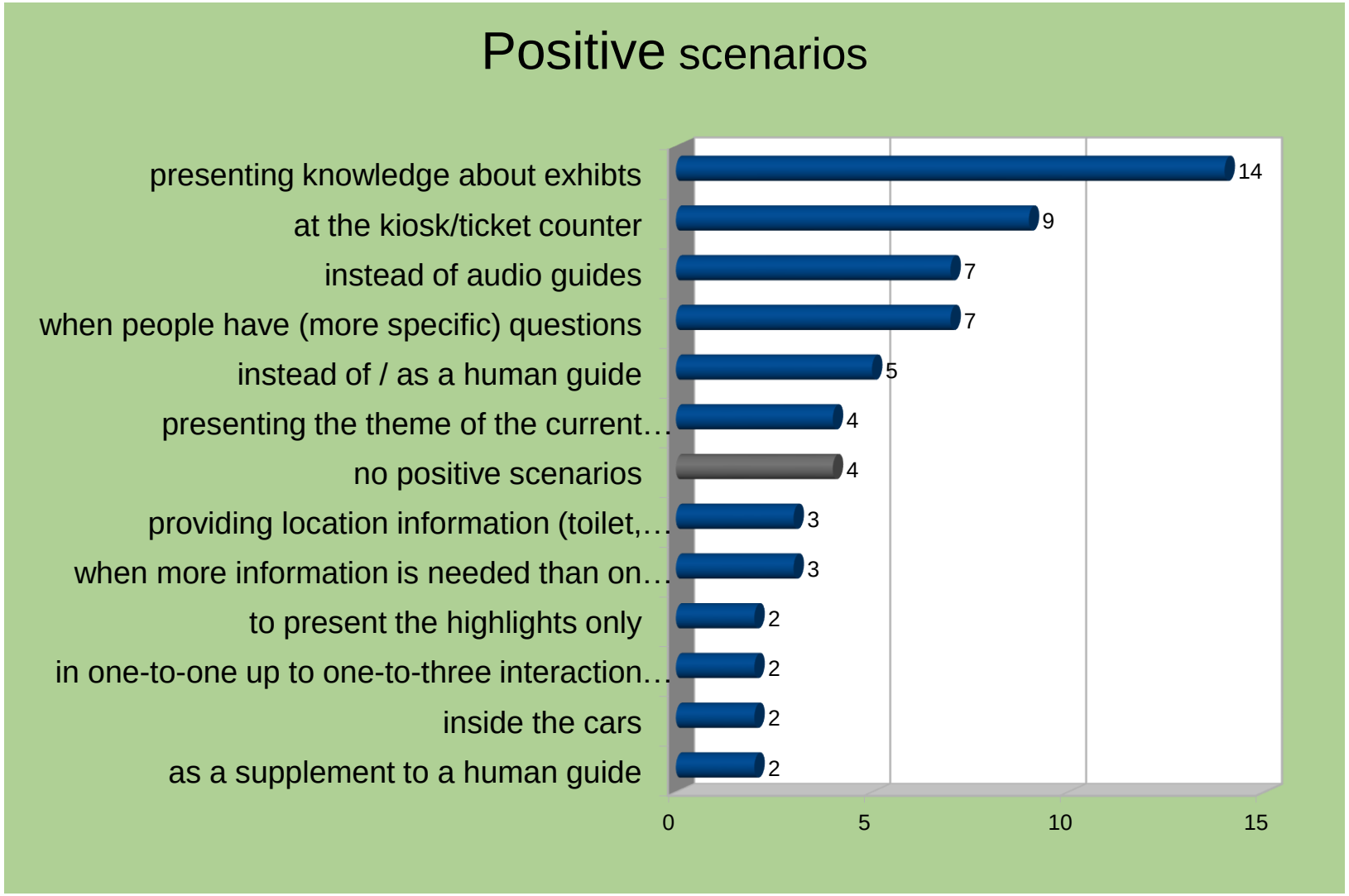
HOCHSCHULE DER MEDIEN

- Many visitors did not come up with any negative scenarios
- Some do not want Andrea instead of an audio guide

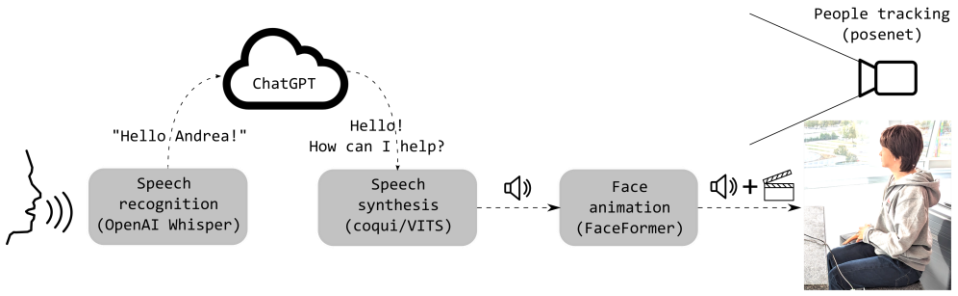


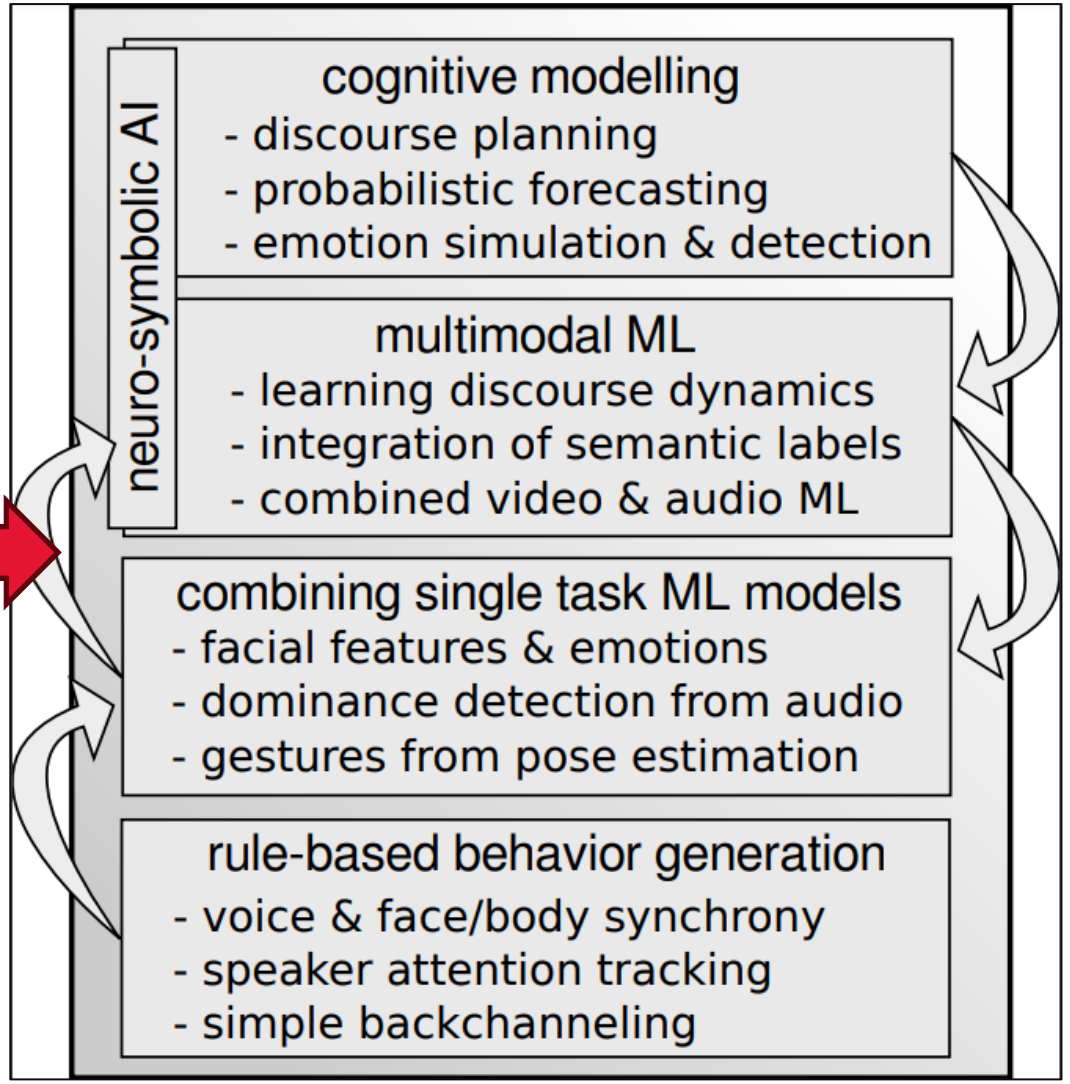
Scenarios that are making sense

- Nearly every third visitor wants to let Andrea present exhibits
- Nine can imagine to let Andrea work at the ticket counter
- Even as a replacement of audio guides is possible for some visitors



What‘s next?


People tracking
(posenet)
ChatGPT
"Hello Andrea!"
Hello!
How can I help?
Speech
recognition
(OpenAI Whisper)
Speech
synthesis
(coqui/VITS)
Face
animation
(FaceFormer)


neuro-symbolic AI
cognitive modelling
- discourse planning
- probabilistic forecasting
- emotion simulation & detection
multimodal ML
- learning discourse dynamics
- integration of semantic labels
- combined video & audio ML
combining single task ML models
- facial features & emotions
- dominance detection from audio
- gestures from pose estimation
rule-based behavior generation
- voice & face/body synchrony
- speaker attention tracking
- simple backchanneling

Cooperation with Japan – ERICA / ATR

Humanoid Lab at Stuttgart Media University

Marcel Heisler
Stuttgart Media University
ML, software framework, facial animation

Johanna Kuch
University of Augsburg
Speech & Gender neutral design

Leon Kiefer
Stuttgart Media University
ML, Vision, Object recognition